

I claim:

Claim 1. A computer system for generating data structures for information retrieval of documents stored in a database, said documents being stored as document-keyword vectors generated from a predetermined keyword list, and said document-keyword vectors forming nodes of a hierarchical structure imposed upon said documents, said computer system comprising:

a neighborhood patch generation subsystem for generating groups of nodes having similarities as determined using a search structure, said neighborhood patch generation subsystem

10 including a subsystem for generating a hierarchical structure upon said document-keyword vectors and a patch defining subsystem for creating patch relationships among said nodes with respect to a metric distance between nodes; and

a cluster estimation subsystem for generating cluster data of said document-keyword vectors using said similarities of patches.

15

Claim 2. The computer system of claim 1, wherein said computer system comprises a confidence determination subsystem for computing inter-patch confidence values between said patches and intra-patch confidence values, and said cluster estimation subsystem selects

20 said patches depending on said inter-patch confidence values to represent clusters of said document-keyword vectors.

Claim 3. The computer system of claim 1, wherein said cluster estimation subsystem

25 estimates sizes of said clusters depending on said intra-patch confidence values.

Claim 4. A method for generating data structures for information retrieval of documents stored in a database, said documents being stored as document-keyword vectors

generated from a predetermined keyword list, and said document-keyword vectors forming nodes of a hierarchical structure imposed upon said documents, said method comprising the steps of:

generating a hierarchical structure upon said document-keyword vectors and storing hierarchy

5 data in an adequate storage area;

generating neighborhood patches of nodes having similarities as determined using levels of the hierarchical structure, and storing said patches in an adequate storage area;

invoking said hierarchy data and said patches to compute inter-patch confidence values between said patches and intra-patch confidence values, and storing said values as

10 corresponding lists in an adequate storage area; and

selecting said patches depending on said inter-patch confidence values and said intra-patch confidence values to represent clusters of said document-keyword vectors.

15 Claim 5. The method according to claim 4 further comprising the step of estimating sizes of said clusters depending on said intra-patch confidence values.

Claim 6. A program for making a computer system execute a method for generating

20 data structures for information retrieval of documents stored in a database, said documents being stored as document-keyword vectors generated from a predetermined keyword list, and said document-keyword vectors forming nodes of a hierarchical structure introduced into said documents, said program making said computer system execute the steps of:

generating a hierarchical structure upon said document-keyword vectors and storing hierarchy

25 data in an adequate storage area;

generating neighborhood patches consisting of nodes having similarities as determined using levels of the hierarchical structure, and storing said patches in an adequate storage area;

invoking said hierarchy data and said patches to compute inter-patch confidence values between said patches and intra-patch confidence values, and storing said values as

corresponding lists in an adequate storage area; and  
selecting said patches depending on said inter-patch confidence values and said intra-patch  
confidence values to represent clusters of said document-keyword vectors.

5

Claim 7. The method according to claim 6, further comprising the step of estimating  
sizes of said clusters depending on said intra-patch confidence values.

10 Claim 8. A computer readable medium storing a program for making a computer  
system execute a method for generating data structures for information retrieval of documents  
stored in a database, said documents being stored as document-keyword vectors generated  
from a predetermined keyword list, and said document-keyword vectors forming nodes of a  
hierarchical structure imposed upon said documents, said program making said computer  
15 system execute the steps of:  
generating a hierarchical structure upon said document-keyword vectors and storing hierarchy  
data in an adequate storage area;  
generating neighborhood patches consisting of nodes having similarities as determined using  
levels of the hierarchical structure, and storing said patch list in an adequate storage area;  
20 invoking said hierarchy data and said patches to compute inter-patch confidence values  
between said patches and intra-patch confidence values, and storing said values as  
corresponding lists in an adequate storage area; and  
selecting said patches depending on said inter-patch confidence values and said intra-patch  
confidence values to represent clusters of said document-keyword vectors.

25

Claim 9. The method according to claim 8, further comprising the step of estimating  
sizes of said clusters depending on said intra-patch confidence values.

Claim 10. An information retrieval system for of documents stored in a database, said documents being stored as document-keyword vectors generated from a predetermined keyword list, and said document-keyword vectors forming nodes of a

5 hierarchical structure imposed upon said documents, said system comprising:  
a neighborhood patch generation subsystem for generating groups of nodes having similarities as determined using a hierarchical structure, said patch generation subsystem including a subsystem for generating a hierarchical structure upon said document-keyword vectors and a patch defining subsystem for creating patch relationships among said nodes with respect to a  
10 metric distance between nodes; and  
a cluster estimation subsystem for generating cluster data of said document-keyword vectors using said similarities of patches; and  
a graphical user interface subsystem for presenting said estimated cluster data on a display means.

15

Claim 11. The computer system of claim 10, wherein said information retrieval system comprises a confidence determination subsystem for computing inter-patch confidence values between said patches and intra-patch confidence values, and said cluster estimation  
20 subsystem selects said patches depending on said inter-patch confidence values to represent clusters of said document-keyword vectors.

Claim 12. The system of claim 10, wherein said cluster estimation subsystem

25 estimates sizes of said clusters depending on said intra-patch confidence values.

Claim 13. The system of claim 10, wherein said system further comprises a user query receiving subsystem for receiving said query and extracting data for information retrieval to

generate a query vector, and an information retrieval subsystem for computing similarities between said document-keyword vectors and said query vector to select said document-keyword vectors.

5

Claim 14. The system of claim 10, wherein said clusters are estimated using said retrieved document-keyword vectors with respect to said user input query.

10 Claim 15. A graphical user interface system for graphically presenting estimated clusters on a display device in response to a user input query, said graphical user interface system comprising:

a database for storing documents;

a computer for generating document-keyword vectors for said documents stored in said

15 database and for estimating clusters of documents in response to said user input query; and a display for displaying on screen said estimated clusters together with confidence relations between said clusters and hierarchical information pertaining to cluster size.

20 Claim 16. The graphical user interface system of claim 15, wherein said computer comprises:

a neighborhood patch generation subsystem for generating groups of nodes having similarities as determined using a search structure, said neighborhood patch generation subsystem including a subsystem for generating a hierarchical structure upon said document-keyword

25 vectors and a patch defining subsystem for creating patch relationships among said nodes with respect to a metric distance between nodes; and

a cluster estimation subsystem for generating cluster data of said document-keyword vectors using said similarities of patches.

**Claim 17.** The graphical user interface system of claim 15, wherein said computer comprises a confidence determination subsystem for computing inter-patch confidence values between said patches and intra-patch confidence values, and said cluster estimation subsystem 5 selects said patches depending on said inter-patch confidence values to represent clusters of said document-keyword vectors and said cluster estimation subsystem estimates sizes of said clusters depending on said intra-patch confidence values.